PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/00	A2	(11) International Publication Number: **WO 98/37180**
		(43) International Publication Date: · 27 August 1998 (27.08.98)

(21) International Application Number: PCT/US98/03367

(22) International Filing Date: 20 February 1998 (20.02.98)

(30) Priority Data:
08/804,266   22 February 1997 (22.02.97)   US

(71) Applicant: ABBOTT LABORATORIES [US/US]; CHAD 0377/AP6D-2, 100 Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(72) Inventors: CHEN, Chih-Ming; 1306 Deer Run Road, Gurnee, IL 60031 (US). MOLLA, Akhteruzzaman; 1249 Vista Drive, Grunee, IL 60031 (US). TRIPATHI, Rakesh, L.; 108 Ironwood Court, Rolling Meadows, IL 60008 (US).

(74) Agents: CASUTO, Dianne et al.; Abbott Laboratories, CHAD 0377/AP6D-2, 100 Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(81) Designated States: CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HCV FUSION PROTEASE AND POLYNUCLEOTIDE ENCODING SAME

(57) Abstract

The present invention provides biologically active fusion proteins of hepatitis C virus NS3 protease and NS4A cofactor protein which are non-autocleavable and polynucleotides encoding same. Expression vectors comprising those polynucleotides and host cells transformed with those polynucleotides are also disclosed. The invention also provides a method for identifying inhibitor compounds of hepatitis C virus NS3 protease using the disclosed fusion proteins.

HCV FUSION PROTEASE AND POLYNUCLEOTIDE ENCODING SAME

Technical Field

The present invention relates in general to recombinant proteins and recombinant
5   polynucleotides encoding such proteins. More particularly, the present invention concerns a
biologically active protease of HCV, to polypeptide analogs thereof and to polynucleotides
encoding the same.

Background

10   Hepatitis C virus (HCV) is a causative agent of posttransfusion non-A, non-B hepatitis
(Choo, Q.L. et al., Science, 244: 259-362 (1989) and Kuo, G. et al., Science, 244: 362-364
(1989)). From analysis of the viral genome and the putative viral proteins encoded in the
genome, HCV is believed to be a member of the family Flavivridae. The HCV genome has a
single open reading frame that encodes a precursor polyprotein of about 3,000 amino acid
15   residues. (Choo, Q.-L., et al., Proc. Natl. Acad. Sci. USA , 88: 2451-2455 (1991)).
Analysis of proteolytic processing has revealed that the polyprotein is composed of at least 10
viral proteins which appear in the following order: NH_2-Core-E1-E2-p7-NS2-NS3-NS4A-
NS4B-NS5A-NS5B-COOH. The Core (nucleocapsid), E1 and E2 (envelope type 1 and type
2) proteins are structural and believed to be processed by host signal peptidases. The "NS"
20   proteins are believed to be non-structural and involved in viral RNA replication. (Steinkuhler,
C., et al., J. Biol. Chem., 271(11): 6367-6373 ((1995).

In HCV, production of mature viral proteins is accomplished by a series of
cotranslational and posttranslational proteolytic processing steps mediated by two virally
encoded proteases. One of these two proteases, designated "NS2/3", is a metalloprotease, and
25   is encoded in the regions from the C-terminal portion of NS2 to the N-terminal one-third of
NS3. The NS2/3 protease cleaves the NS2/NS3 junction of native HCV polyprotein in cis.
The second protease, designated "NS3", is a serine-type protease encoded in the N-terminal
one-third of NS3. The NS3 protease cleaves at all known NS junctions located downstream
from the NS3 region, namely, at the NS3/4A, NS4A/4B, NS4B/5A and NS5A/5B junction
30   sites. (Sitoh, S., et al., J. Virol., 69(7): 4255-4260 (1995).

NS3 protease processing at the NS3/4A junction appears to take place exclusively as an
intramolecular or cotranslational reaction (in cis). In contrast, cleavage at the other sites can
also be mediated intermolecularly or posttranslationally (i.e. in trans) (Steinkuhler, C., et al.,
op. cit.). Furthermore, cleavage by the NS3 protease at the NS4B/NS5A junction requires an
35   additional cofactor protein encoded by NS4A (see Failla, C. et al., J. Virol., 68(6): 3753-
3760 (1994); Lin, C. et al., J. Virol., 68(12): 8147-8157 (1994); Tanji, Y. et al., J. Virol.,
69(3): 1575-1581 (1994); Bartenschlager, R. et al., J. Virol., 69(1): 198-205 (1995)).
NS4A may act by stabilizing the active conformation of the NS3 protease domain and recruiting

2

NS3 to the membranes, where presumable proteolytic processing takes place (Hijikata, M. *et al.*, *Proc. Natl. Acad. Sci. USA* **90**: 10773-10777 (1993)). However, the actual mechanism by which NS4A and NS3 interact to effect cleavage at the NS4B/5A junction is unknown.

Because the NS3 protease is likely to be an essential enzyme for viral growth, it has become a target for the development of anti-HCV drugs. Toward this end, assays have been developed to screen for drugs which inhibit NS3 protease activity. In such assays, it is generally necessary to provide at least a cleavable substrate, an NS3 protease capable of cleaving the substrate and a compound of interest. However, when the cleavable portion of the substrate is an NS4B/5A junction, it is also necessary to provide a sufficient quantity of NS4A cofactor protein to bring about efficient cleavage. Even when other NS junctions form the cleavage site in a substrate (i.e. NS4A/4B or NS5A/5B), addition of NS4A cofactor protein is desirable since it also renders cleavage more efficient (Failla, C. *et al.*, and Lin, C. *et al.*, *op. cit.*)

One problem that arises in effecting these assays, is in obtaining sufficient quantities of NS3 protease and NS4A cofactor protein to carry out screening assays on a large-scale basis. A second complication that arises is in having to make and/or purify the two proteins separately and then empirically determine the proper proportions of each protein to add to the assay in order to achieve efficient cleavage. This second problem is particularly difficult to overcome, since biologically active NS3 protease is autocleavable at the NS3/4A junction, and therefore self-cleaves itself from NS4A during the purification process. Thus there is a need for a simple, rapid, and cost effective means of generating purified NS3 protease and NS4A cofactor protein in large quantities. There is also a need for a single polypeptide of NS3 protease and NS4A cofactor protein that is easily purified and biologically active and which eliminates the need to reconstitute both proteins in proper proportions to obtain efficient substrate cleavage.

Summary of the Invention

In one aspect, the present invention provides an isolated or purified polynucleotide, comprising a nucleotide sequence (A) having a nucleotide sequence (B) or fragments thereof which encode hepatitis C virus NS3 protease and a nucleotide sequence (C) or fragments thereof which encode NS4A cofactor protein, wherein the nucleotide sequence (A) produces, upon expression, a non-autocleavable fusion protein of hepatitis C virus NS3 protease and hepatitis C virus NS4 cofactor protein which is biologically active. In a preferred embodiment, the nucleotide sequence (B) is located upstream from nucleotide sequence (C). Furthermore, the nucleotide (A) encodes a biologically active fusion protein which is capable of cleaving at least SEQ ID NO:15. In one embodiment, the nucleotide sequence (B) encodes a biologically active domain of NS3 protease. In a more preferred embodiment, the nucleotide sequence (B)

comprises from about nucleotide position 1 to about nucleotide position 543 of SEQ ID NO:1. In another embodiment, the nucleotide sequence (C) encodes a biologically active domain of NS4A cofactor protein which more preferably, comprises from about nucleotide position 1957 to about nucleotide position 1995 of SEQ ID NO:1. In a most preferred embodiment, the nucleotide sequence (A) has the sequence of SEQ ID NO:3.

In another embodiment, a polynucleotide of the present invention is contained in an expression vector. The expression vector preferably further comprises an enhancer-promoter operatively linked to the polynucleotide. A preferred expression vector is pGEX. In a more preferred embodiment, the pGEX vector comprises the polynucleotide of SEQ ID NO:3.

The present invention still further provides for a host cell transformed with an expression vector of this invention. The host cell may be a eukaryotic or prokaryotic cell. Preferably, the host cell is *E. coli*.

The present invention also provides a biologically active fusion polypeptide comprising hepatitis C virus NS3 protease and hepatitis C virus NS4A cofactor protein which is non-autocleavable. The fusion protein is capable of cleaving at least SEQ ID NO:16 and preferably, also cleaves a substrate comprising SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9. In a preferred embodiment, the fusion protein has SEQ ID NO:4.

In yet another embodiment, the present invention provides a method for identifying an inhibitor compound of hepatitis C virus NS3 protease comprising the steps of (a) providing a reaction mixture having (i) a substrate wherein the substrate is capable of being cleaved by a hepatitis C virus NS3 protease acting alone or in combination with a hepatitis C virus NS4A cofactor protein, (ii) a non-autocleavable fusion protein of hepatitis C virus NS3 protease and hepatitis C virus NS4A cofactor protein which is biologically active and (iii) a compound of interest; (b) incubating said reaction mixture; and (c) determining the extent of cleavage of said substrate in said reaction mixture. Preferably in the method, the fusion protein has SEQ ID NO:3.

Brief Description of the Drawings

FIG. 1 shows a partial polynucleotide sequence of an HCV genome, strain H (SEQ ID NO:1) and is intended to represent both the sense strand (which is shown) and its complementary strand. Standard one letter codes for the amino acids appear beneath their respective nucleic acid codons.

FIG. 2 shows a polynucleotide sequence (SEQ ID NO:3) which encodes an NS3/4A fusion protein of the present invention. This particular sequence represents the sense sequence of SEQ ID NO:1 from about nucleotide position 1 to about nucleotide position 612 and from

4

about nucleotide position 1894 to about nucleotide position 2055.

FIG. 3 shows the polypeptide sequence (SEQ ID NO:4) encoded from SEQ ID NO:2.

FIG. 4 shows a graph of the results of a kinetics assay performed as described in Example 3. In the graph, the closed circles, plus sign symbols, "x" symbols and open circles represent fluorescence points obtained from assays performed in the presence of pT-3 fusion protein, glutathione S transferase (GST), GST coupled to cytomegalovirus (CMV) protease, and no enzyme, respectively.

FIG. 5 depicts the HPLC analysis of cleavage products after incubation of a purified GST-NS3/4A fusion protein with a cleavable substrate (i.e. SEQ ID NO:16). The assay was performed under conditions described in Example 3 (Total Cleavage Assay). Aliquots from the total cleavage assay were withdrawn at the time points indicated to the left of the HPLC tracings. Time points indicated below the tracings show the peak retention times. The dotted lines represent 470 nm absorption and the solid lines represent the fluorescence tracing with excitation at 355nm and emission at 490 nm.

FIG. 6 schematically shows the T3 and NS3 series of fusion constructs of NS3/4A [FIG 6(a)] and NS3 [FIG 6(b)] fused downstream of maltose binding protein and protease cleavage sites in pMAL vectors.

Detailed Description

I. The Invention

The present invention provides polynucleotide sequences which encode a fusion protein of hepatitis C virus (hereinafter HCV) NS3 protease and hepatitis C virus NS4A cofactor protein. Such sequences may include: the incorporation of codons "preferred" for expression by desired non-mammalian hosts, the provision of sites for cleavage by restriction endonuclease enzymes; and the provision of additional initial, terminal or intermediate DNA sequences which facilitate construction of readily expressed vectors.

In another embodiment, the present invention provides a recombinant fusion protein of hepatitis C virus which is biologically active. Furthermore, the invention also includes expression vectors for high level expression and easy purification and host cells transformed with such vectors.

II. Definitions

5

For the purposes of the present invention as disclosed and claimed herein, the following terms are defined.

The term "NS3 protease" as used herein refers to a serine-type protease encoded by HCV which is capable, either alone or in combination with NS4A cofactor protein (described below), of cleaving a substrate having an HCV non-structural (NS) cleavage junction (defined below). The term NS3 protease is intended to encompass protease analogs (defined below) provided such analogs also possess the ability to cleave an HCV NS cleavage junction as described below.

The term "NS4A cofactor" or "NS4A cofactor protein" as used herein refers to a protein encoded by HCV which acts in combination with NS3 protease, to effect cleavage of a substrate having an HCV non-structural (NS) cleavage junction as described below. Although NS4A cofactor is believed to effect cleavage by stabilizing the NS3 protease and/or recruiting NS3 protease to the membrane, the actual mechanism by which NS4A cofactor acts "in combination" with NS3 protease is unknown. The term NS4A cofactor is also intended to include protein analogs of NS4A cofactor provided those analogs possess the ability to act in combination with NS3 protease to effect cleavage of a cleavage junction.

The term "polypeptide" as used herein refers to a molecular chain of amino acids and does not refer to a specific length of the product. Thus, peptides, oligopeptides and proteins are included within the definition of polypeptide. Hepatitis C virus NS3 protease and NS4A cofactor protein are representative examples of polypeptides. This term is also intended to refer to post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like.

The term "fusion protein" as used herein refers to a polypeptide comprising an amino acid sequence drawn from two or more individual proteins. A fusion protein is formed by the expression of a polynucleotide in which at least two coding sequences have been joined together such that their reading frames are in frame. Examples of fusion proteins of the present invention include a polypeptide comprising NS3 protease joined to NS4A cofactor protein or an NS3/4A fusion protein further joined to a biological tag. Such fusion proteins may or may not be capable of being cleaved into the separate proteins from which they are derived.

The term "cleavage junction" or "non-structural cleavage junction" as used herein refers to a polypeptide comprising a continguous sequence of amino acids having the formula X_6-X_5-X_4-X_3-X_2-X_1-X_1' (SEQ ID NO:5) wherein X_6 represents D or E, X_1 represents T or C, X_1' represents A or S and X_2, X_3, X_4, and X_5 represent any amino acid. Such a cleavage junction is further defined as one which NS3 protease alone or in combination with NS4A cofactor protein can cleave. As determined by Steinkühler et al., (J. Biol. Chem., 271(11): 6367-6373 (1995)), the amino acid sequence "D/E-X_5-X_4-X_3-X_2-C-A/S" represents a consensus sequence for all NS3 trans cleavage sites (i.e. sites which are cleaved by NS3 protease alone or

6

in combination with NS4A via an intermolecular reaction). In this consensus sequence, each single letter (i.e. D, E, C, A and S) represents aspartic acid, glutamic acid, cysteine, alanine and serine respectively; the slash symbol "/" designates the word "or" and X_2-X_5 represent any amino acids.. The consensus sequence for *in cis* or (intramolecular) cleavage differs slightly

5   from the other in having a T (threonine) residue present instead of C at the X_1 position.

Also contained within the *trans* and *cis* consensus sequence is a scissile bond or point of actual cleavage. In accordance with the nomenclature of Berger and Schechter (*Philos. Trans. R. Soc. Lond. B* 257: 249-264 (1970)) and as used throughout this specification, a newly generated carboxy terminal amino acid, created after cleavage of a peptide bond, is

10   designated as P1 and is preceded by a P2 residue which is preceded by a P3 residue etc; a newly generated amino terminus is designated P1' and is followed by P2', P3', P4' etc. In the *trans* and *cis* consensus sequences described above, C and T are P1 residues, A and S are P1' residues X_2-X_5 are residues P2, P3, P4 and P5 and D or E is the P6 residue. Similarly in SEQ ID NO:4, X_1 represents a P1 residue, $X_{1'}$ a P1' residue, X_6 a P6 residue etc.

15   The term "cleavable substrate" as used herein refers to a polypeptide comprising at least the cleavable junction of SEQ ID NO:5. Examples of cleavable substrates include a native HCV polyprotein and fragments thereof. Preferred cleavable substrates include polypeptides comprising SEQ ID NO:5 wherein SEQ ID NO:5 has the sequence of a native HCV NS junction selected from the group consisting of NS3/4A = DLEVVTS (SEQ ID NO:6),

20   NS4A/4B = DEMEECS (SEQ ID NO:7), NS4B/5A = ECTTPCS (SEQ ID NO:8), and NS5A/5B = EDVVCCS (SEQ ID NO:9). Even more preferred cleavable substrates comprise sequences selected from the group consisting of DLEVVTSTWVL (SEQ ID NO:10), DEMEECSQHLP (SEQ ID NO:11), ECTTPCSGSWL (SEQ ID NO:12), and EDVVCCSMSYT (SEQ ID NO:13). Other preferred cleavable substrates include E-A-G-D-D-

25   I-V-P-C-S-M-S-Y-T-W-T-G-A (SEQ ID NO: 14, see Shimizu *et al.*, *Virology* 70(1): 127-132 (1996)) and E-D-V-V-C-C-S-M-S-Y (SEQ ID NO:15, see Steinkuhler *et al.*, *J. Virology* 70(10): 6694-6700 (1996)). Cleavable substrates may be generated in any manner well known to those of ordinary skill in the art, such as by synthetic means or by proteolytic digestion of a native HCV polyprotein.

30   Cleavable substrates need not be of any specific length but preferably provide detectable cleavage products upon cleavage of the substrate. For example, cleavage products may be assayed by western blot or if the cleavage substrate has been radiolabled, by autoradiography techniques. Alternatively, one or more ends may be labeled with an enzyme so as to permit visualization of the protein products. It is presently preferred to employ small peptide p-

35   nitrophenyl esters or methylcoumarins, as cleavage may then be followed by spectrophotometric or fluorescent assays. For example, following the method described by E.D. Matayoshi *et al.*, (*Science*, 247: 231-235 (1990)) one may attach a fluorescent label to

7

one end of the substrate and a quenching molecule to the other end; cleavage is then determined by measuring the resulting increase in fluorescence. An example of such a cleavable substrate is Ac-G-E(EDANS)-(ethylene glycol linker)-E-D-V-V-A-C-S-M-S-Y-(ethylene glycol linker)-K(Dabycl)-G-NH$_2$ (SEQ ID NO:16).

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, which is separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "polynucleotide" as used herein means a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, the term includes double- and single-stranded DNA, as well as double- and single-stranded RNA. It also includes modifications, either by methylation and/or by capping, and unmodified forms of the polynucleotide.

"Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and more preferably, less than about 90% of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for example, disruption of the cell containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity chromatography and sedimentation according to density. Thus, "purified polypeptide" means a polypeptide of interest or fragment thereof which is essentially free, that is, contains less than about 50%, preferably less than about 70%, and more preferably, less than about 90% of cellular components with which the polypeptide of interest is naturally associated. Methods for purifying are well known to those of ordinary skill in the art.

The term "open reading frame" or "ORF" refers to a region of a polynucleotide sequence which is not interrupted by any stop codons; this region may represent a portion of a coding sequence or a total coding sequence.

The term "recombinant protein" or "recombinant polypeptide" as used herein refers to at least a polypeptide of genomic, semisynthetic or synthetic origin which by virtue of its origin or manipulation is not associated with all or a portion of the polypeptide with which it is associated in nature or in the form of a library and/or is linked to a polypeptide other than that to which it is linked in nature. A recombinant polypeptide may be translated from a designated sequence of HCV or HCV genome. However, it also may be generated in other ways, such as

8

by chemical synthesis or via expression in a recombinant expression system, or by isolation from a mutated HCV.

The term "recombinant host cells", "host cells", "cells", "cell lines", "cell cultures" and other such terms denoting microorganisms or higher eucaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the original progeny of the original cell which has been transfected.

The term "replicon" as used herein means any genetic element, such as a plasmid, a chromosome, a virus, that behaves as an autonomous unit of polynucleotide replication within a cell. Otherwise stated, a replicon is a genetic element which is capable of replication under its own control.

The term "vector" as used herein refers to a replicon in which another polynucleotide segment is attached, such as to bring about the replication and/or expression of the attached segment.

The term "control sequence" as used herein, refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism. In prokaryotes, such control sequences generally include promoters, ribosomal binding sites and terminators; in eukaryotes, such control sequences generally include promoters, terminators and in some instances, enhancers. Thus the term "control sequence" is intended to include at a minimum all components whose presence is necessary for expression, and also may include additional components whose presence is advantageous, for example, leader sequences.

The term "operatively linked" refers to a situation in which the components described are are in a relationship permitting them to function in their intended manner. Thus, for example, a control sequence "operatively linked" to a coding sequence is ligated in such a manner that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The term "coding sequence" as used herein refers to a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to, mRNA, cDNA and recombinant polypeptide sequences.

The term "transformation" refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion. For example, direct uptake, transduction, or f-mating are included. The exogenous polynucleotide may be maintained as a non-integrated vector such as for example, a plasmid, or alternativelym may be integrated into

the host genome.

II. NS3/NS4A Polynucleotides

In one aspect, the present invention provides an isolated or purified polynucleotide
comprising a nucleotide sequence (A) which encodes a fusion protein of NS3 protease and
NS4A cofactor protein from hepatitis C virus (HCV). Hereinafter, the fusion protein
expressed from such a nucleotide will be referred to as "NS3/4A fusion protein".

FIG. 1 (SEQ ID NO:1) shows a partial polynucleotide sequence of an HCV genome
(strain H), specifically, the polynucleotide sequence encoding native NS3 protease and NS4A
cofactor protein and is intended to represent both the sense strand (as shown) and its
complement. The polypeptide encoded therefrom (SEQ ID NO:2) is shown below with
standard one letter codes for the amino acids appearing beneath their respective nucleic acid
codons. In SEQ ID NO:1, the nucleotide sequence which encodes NS3 protease is located
from about nucleotide position 1 to about nucleotide position 1893. The smallest portion of
nucleotide sequence known to encode a biologically active NS3 protease is from about
nucleotide position 1 to about nucleotide position 546. Also shown in SEQ ID NO:1 is a
nucleotide sequence of NS4A cofactor protein, which is located from about nucleotide position
1894 to about nucleotide position 2055. The smallest portion of nucleotide sequence known to
encode a biologically active NS4A cofactor protein is from about nucleotide position 1954 to
about nucleotide position 1995. As can be seen from SEQ ID NO:1, the polynucleotide
contains a continuous open reading frame.

A polynucleotide sequence of the present invention comprises a nucleotide sequence (A)
derived from SEQ ID NO:1 having a nucleotide sequence (B) which encodes an NS3 protease
and a nucleotide sequence (C) which encodes an NS4A cofactor protein in a continuous
translational open reading frame. In a preferred embodiment, the polynucleotide comprises a
nucleotide sequence having the sense sequence of SEQ ID NO:1 from about nucleotide position
1 to about nucleotide position 612 and about nucleotide position 1894 to about nucleotide
position 2055. Such a preferred polynucleotide is shown in FIG. 2 (SEQ ID NO:3).
Furthermore, in a most preferred embodiment, the sequence which encodes the NS3 protease is
located upstream (in front of) the sequence which encodes the NS4A cofactor protein (see again
SEQ ID NO:3). An even more preferred polynucleotide is a DNA molecule. In another
embodiment, the polynucleotide is an RNA molecule.

A polynucleotide sequence of the present invention is further defined as one which
encodes a non-autocleavable fusion protein of NS3 protease and NS4A cofactor protein. Such
a polynucleotide is one which lacks the nucleotide sequence that encodes SEQ ID NO:5;
accordingly, the fusion protein encoded from the polynucleotide will not itself contain a
cleavable junction. As can be seen by a comparison of SEQ ID NO:1 and SEQ ID NO:3, SEQ

ID NO:3 lacks the nucleotide sequence encoding the terminal portion of native NS3 protease
(i.e. it is missing the nucleotides from position 613 to position 1893 of SEQ ID NO:1),
including that particular sequence which encodes SEQ ID NO:5 (i.e. from nucleotide position
1876 to nucleotide position 1893 of SEQ ID NO:1). Thus, the NS3 portion of the polypeptide
5 . encoded from SEQ ID NO:3 is unable to cleave itself from the fusion protein (shown in FIG.
3, SEQ ID NO:4).

It is to be noted that the manner of making such a fusion protein is not critical to the
practice of the invention. For example, if a non-autocleavable fusion protein is to be generated
from an autocleavable sequence (such as an HCV genome or portion thereof), one or more of
10 the SEQ ID NO:5 nucleotides contained within that genomic sequence may be eliminated either
by deletion, mutation or addition of sequence (so as to disrupt SEQ ID NO:5). The only
requirements are that the resulting nucleotide sequence encode a non-autocleavable junction and
retain an open reading frame between the coding regions of NS3 and NS4A so that the
polypeptide encoded therefrom will be biologically active.

15   For the purpose of measuring biological activity only, a polypeptide of the present
invention must be shown to cleave at least a cleavable substrate SEQ ID NO:16 when tested as
described in Example 3 below. It is to be understood however, that such a polypeptide may
also cleave other cleavable substrates, both natural and synthetic. For example, a biologically
active protease encoded from a polynucleotide of the present invention may also possess the
20 ability to cleave a native HCV genome or fragments thereof or other cleavable substrates as
described herein.

The present invention also contemplates shorter and longer polynucleotide sequences
(other than that shown in SEQ ID NO:3) which encode an NS3/NS4A fusion protein provided
that the fusion protein possesses the characteristics of being non-autocleavable and biologically
25 active. For example, the present invention also contemplates polynucleotide sequences which
encode the smallest proteolytic domain of an NS3 protease (i.e. from about nucleotide position
1 to about nucleotide position 543 of SEQ ID NO:1) or the smallest proteolytic domain of an
NS4A cofactor protein (i.e. from about nucleotide position 1957 to about nucleotide position
1995 of SEQ ID NO:1) or both provided that such domains form a fusion protein that
30 possesses biological activity as defined above. Thus, the polynucleotides contemplated by the
present invention include those which contain at least active domains of NS3 protease and
NS4A cofactor protein. In addition, when constructing such polynucleotide sequences, the
sequences must retain the characteristics of having a single open reading frame and of encoding
a non-autocleavable fusion protein. Standard molecular biology techniques are used for
35 generating such polynucleotides and are well known to those of ordinary skill in the art (see for
example, Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, (Cold
Spring Harbor, N.Y., 1989).

11

The present invention also contemplates analogous DNA sequences which hybridize under stringent hybridization conditions to the DNA sequences set forth above. Stringent hybridization conditions are well known in the art and define a degree of sequence identity greater than about 80% and more preferably, greater than about 90%. The modifier "analogous" also refers to those nucleotide sequences that encode polypeptides having only conservative differences and which retain the conventional characteristics and activities of an NS3/NS4A fusion protein; eg. cleaving SEQ ID NO:16. The present invention also contemplates naturally occurring allelic variations and mutations of the DNA sequences set forth above so long as those variations and mutations code, on expression, for an NS3/4A fusion protein of this invention as set forth hereinafter.

As is well known in the art, because of the degeneracy of the genetic code, there are numerous other DNA and RNA molecules that can code for the same polypeptide as those of a particular sequence. The present invention, therefore, contemplates those other DNA and RNA molecules which, on expression, encode for the polypeptide of NS3/4A fusion protein or fragments thereof. Having identified the amino acid residue sequence encoded by an NS3/4A polynucleotide, and with knowledge of all triplet codons for each particular amino acid residue, it is possible to describe all such encoding RNA and DNA sequences. DNA and RNA molecules other than those specifically disclosed herein and, which molecules are characterized simply by a change in a codon for a particular amino acid are within the scope of this invention.

A polynucleotide of the present invention can also be an RNA molecule. A RNA molecule contemplated by the present invention is complementary to or hybridizes under stringent conditions to any of the DNA sequences set forth above. Exemplary and preferred RNA molecules are mRNA molecules that encode an NS3/NS4A fusion protein of this invention.

II. HCV NS3 Protease/NS4A Cofactor Fusion Protein

In another aspect, the present invention provides a fusion protein of NS3 protease and NS4A cofactor of HCV. An NS3/NS4A fusion protein of the present invention is a polypeptide of from about 194 amino acid residues which has the ability to cleave at least SEQ ID NO:16 when tested as described in Example 3 below. Such an NS3/4A fusion protein may also have the ability to cleave other cleavable substrates including but not limited to a native HCV genome and fragments thereof. Furthermore, an NS3/4A fusion protein of the present invention is non-autocleavable, meaning that the fusion protein itself SEQ ID NO:16. The amino acid sequence of an exemplary NS3/4A fusion protein is set forth in FIG. 3 (SEQ ID NO:4).

The present invention also contemplates amino acid residue sequences that are substantially duplicative of the sequences set forth herein such that those sequences

12

demonstrate like biological activity to disclosed sequences. Such contemplated sequences include those sequences characterized by a minimal change in amino acid residue sequence or type (e.g., conservatively substituted sequences) which insubstantial change does not alter the fundamental nature and biological activity of an NS3/4A fusion protein.

5   It is well known in the art that modifications and changes can be made in the structure of a polypeptide without substantially altering the biological function of that peptide. For example, certain amino acids can be substituted for other amino acids in a given polypeptide without any appreciable loss of function. In making such changes, substitutions of like amino acid residues can be made on the basis of relative similarity of side-chain substituents, for

10   example, their size, charge, hydrophobicity, hydrophilicity, and the like.

As detailed in United States Patent No. 4,554,101, incorporated herein by reference, the following hydrophilicity values have been assigned to amino acid residues: Arg (+3.0); Lys (+3.0); Asp (+3.0); Glu (+3.0); Ser (+0.3); Asn (+0.2); Gln (+0.2); Gly (0); Pro (-0.5); Thr (-0.4); Ala (-0.5); His (-0.5); Cys (-1.0); Met (-1.3); Val (-1.5); Leu (-1.8); Ile (-1.8); Tyr

15   (-2.3); Phe (-2.5); and Trp (-3.4). It is understood that an amino acid residue can be substituted for another having a similar hydrophilicity value (e.g., within a value of plus or minus 2.0) and still obtain a biologically equivalent polypeptide.

In a similar manner, substitutions can be made on the basis of similarity in hydropathic index. Each amino acid residue has been assigned a hydropathic index on the basis of its

20   hydrophobicity and charge characteristics. Those hydropathic index values are: Ile (+4.5); Val (+4.2); Leu (+3.8); Phe (+2.8); Cys (+2.5); Met (+1.9); Ala (+1.8); Gly (-0.4); Thr (-0.7); Ser (-0.8); Trp (-0.9); Tyr (-1.3); Pro (-1.6); His (-3.2); Glu (-3.5); Gln (-3.5); Asp (-3.5); Asn (-3.5); Lys (-3.9); and Arg (-4.5). In making a substitution based on the hydropathic index, a value of within plus or minus 2.0 is preferred.

25

III.   Method of Making an NS3/4A Fusion Protein

In another aspect the present invention provides a process for making a polynucleotide NS3/4A fusion protein. In accordance with that process, a suitable host cell is transformed with a polynucleotide of the present invention. The transformed cell is maintained for a period

30   of time sufficient for expression of the NS3/4A fusion protein; the fusion protein is then recovered.

The polynucleotide which encodes NS3 protease and/or NS4A cofactor can be obtained in varous ways. For example, the HCV nucleic acid can be isolated and cloned from viral particles obtained from individuals infected with the virus. The gene encoding NS3 protease

35   can also be obtained using the plasmid disclosed in Grakoui, A. et al., J. Virology, 67(3): 1385-1395 (1993)). Alternatively, the polynucleotide of the invention can be chemically synthesized by means well known in the art. (See for example, Matteucci, et al., J. Am.

13

Chem. Soc., **103**: 3185 (1981) and B.R. Glick and Pasternak, *Molecular Biotechnology*, ASM Press, Washington, D.C. pages 55-63 (1994)). Furthermore, the HCV genome has been disclosed in PCT International Application WO 89/04669 and is available from the American Type Culture collection (ATCC), 12301 Parklawn Drive, Rockville, MD under Accession No. 40394.

   a.   Hosts and Expression Systems (Control Sequences and Vectors)

   Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotics include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from the plasmid pBR322 which contains operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. There markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the beta-lactamase (penicillinase), lactose promoter system (Chang et al., *Nature* 198:1056 (1977)), the tryptophan promoter system (reported by Goeddel et al., *Nucleic Acid Res.* 8: 4057 (1980)) and the lambda-derived P1 promoter and N gene ribosome binding site (Shimatake *et al.*, *Nature* 292: 128 (1081) and the hybrid Tac promoter (De Boer *et al.*, *Proc. Natl. Acad. Sci. USA* 292: 128 (1983)) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with *E. coli*; however, other prokaryotic hosts such as strains of *Bacillus* or *Pseudomonas* may be used if desired, with corresponding control sequences.

   Eukaryotic hosts include yeast, mammalian and insect cells in culture systems. *Saccharomyces cerevisiae* and *Saccharomyces carlsbergensis* are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring protrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (as described by Broach *et al.*, *Meth. Enz.* **101**: 307 (1983), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes, including the promoter for 3 phosphophycerate kinase. See, for example, Hess *et al.*, *J. Adv. Enzyme Reg.* 7:149 (1968), Holland *et al.*, *Biochemistry* 17:4900 (1978) and Hitzeman, *J. Biol. Chem.* 255: 2073 (1980). Terminators also may be included, such as those derived from the enolase gene as reported by Holland, *J. Biol. Chem.* 256: 1385 (1981). It is contemplated that particularly useful control systems are those which comprise the

glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequences from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are not naturally associated in the wild-type organism.

Mammalian cell lines available as hosts for expression are known in the art and may include many immortalized cell lines which are available from the American Type Culture Collection. These include HeLa cells, Chinese hamster overy (CHO) cells, baby hamster kidney (BHK) cells, and the like. Suitable promoters for mammalian cells also are known in the art and include viral promoters such as that from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus (ADV), bovine papilloma virus (BPV), and cytomegalovirus (CMV). Mammalian cells also may require terminator sequences and poly A addition sequences; enhancer sequences which increase expression also may be included as well as sequences which cause amplification of a gene. Such sequences are well known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences in a host genome. An example of a mammalian expression system for HCV is described in U.S. Patent Application Serial No. 07/830,024, filed January 31, 1992.

Insect cell lines are also available as hosts and are well known to those of ordinary skill in the art. Cloning vehicles such as baculovirus may be used in such cell lines.

The present invention also comtemplates the use of expression vectors which facilitate purification of a desired polypeptide. For example, a polynucleotide encoding the desired fusion protein may be cloned into an expression vector which, when expressed, produces the fusion protein linked to a chemical or biological tag. A tag may be any chemical or biological compound or fragment thereof capable of binding to a specific substrate or receptor. Thus, tags serve to facilitate purfication of a tagged fusion product via specific binding of the tag portion to its receptor or substrate. Preferably the tag is linked to the polypeptide in a manner that permits it to be cleaved from the polypeptide after purification, without affecting the activity of the polypeptide. In illustration, a polynucleotide of the present invention may be cloned into a pGEX vector (Pharmacia Biotech. Inc., Piscataway, New Jersey) and placed in a suitable host; on expression, a fusion protein of NS3/NS4A and glutathione S-transferase (GST) is produced. The fusion protein is then purified by affinity chromatography using glutathione sepharose 4B (which binds to the GST portion of the fusion product). The NS3/4A fusion protein is then cleaved from the GST tag using a site-specific protease whose recognition sequence is located upstream from the NS3/4A fusion protein. Other affinity tags may also be used and linked to either end of the desired protein (i.e. either amino or carboxyl terminus). For example, tags such as 6-His (available from Novagen, Madison, WI), hexa-Arg (see G.

15

Stempfer et al., Nature Biotechnology 14: 481-484 (1996)), FLAG (available from VWR, Chicago, IL), maltose binding protein (MBP, see Kellerman and Ference, Methods in Enzymology 90: 459-463, 1992) and thioredoxin (Trx, see La Vallie, et al., Bio/Technology 11: 187-193, 1993) may also be used and are well known to routineers.

5   b. Transformations

Means for transforming host cells in a manner such that those cells produce recombinant polypeptides are well known in the art. Such methods include direct uptake of a polynucleotide, packaging a polynucleotide in a virus, and transducing a host cell with a virus. The transformation procedures selected depends upon the host to be transformed. For
10   example, bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride. Cohen, Proc. Natl. Acad. Sci. USA 69: 2110 (1972). Yeast transformation by direct uptake may be conducted using the calcium phosphate precipitation method of Graham et al., Virology 52: 526 (1978) or modification thereof.

c. Vector Construction

15   Vector construction employs methods known in the art. Generally, site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. Usually, about 1 microgram (μg) of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 μl of buffer solution by incubation at 37°C for 1 to 2 hours. After incubation with the
20   restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis methods, according to methods known by the routine practitioner.

Ligations are performed using standard buffer and temperature conditions using T4 DNA ligase and ATP. Sticky end ligations require less ATP and less ligase than blunt end
25   ligations. When vector fragments are used as part of a ligation mixture, the vector fragment often is treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector. Alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation. Ligation mixtures are transformed into suitable cloning hosts such as E coli and successful transformants selected by
30   methods including antibiotic resistance, and then screened for the correct construct.

Uses

An NS3/4A fusion protein of the present invention has numerous uses. By way of example, such a polypeptide can be used in large or small scale in vitro assays for identifying
35   compounds that inhibit the activity of the fusion protein. For example, a fusion protein of the present invention may be incubated with a cleavable substrate and a compound of interest. In such an assay, compounds which prevent the formation of cleavage products are potential

16

inhibitors of protease activity.

An NS3/4A fusion protein can also be used to design compounds that interact with and inhibit the fusion protein. For example, the fusion protein may be used for structural studies by NMR or X-ray diffraction, thereby facilitating drug design.

The invention will be better understood in connection with the following examples, which are intended as an illustration of and not a limitation upon the scope of the invention. Both below and throughout the specification, it is intended that citations to the literature are expressly incorporated by reference.

Example 1: Cloning of plasmids pT-1 and pT-3

Plasmid pBRTM/HCV 1-3011 containing the gene encoding the full length HCV-H polyprotein (amino acids 1-3011) was purchased from Dr. Charles Rice of Washington University School of Medicine. Restriction enzymes were purchased from commercial suppliers such as Boehringer Mannheim (Indianapolis, IN), GIBCO BRL (Gaithersburg, MD) and New England Biolab (Beverly, MA) unless otherwise indicated. Chemicals were purchased from Sigma Chemical Co. (St. Louis, MO).

A. Construction of plasmid pT-1

Plasmid pBRTM/HCV was digested with Kas I and the 2.4 kb fragment (i.e. nucleotides (nt) 7606-10034) was isolated by gel elution. The fragment was treated with Klenow and ligated into vector pHIL-S1 (Invitrogen,San Diego, CA) which had been cut with Sma I and dephosphorylated with calf intestinal alkaline phosphatase (CIAP) in order to generate plasmid pRLT-2. Plasmid pRLT-2 contained the entire open reading frames (ORFs) of both NS3 and NS4A and part of the ORF of NS4B. Plasmid pRLT-3 was then generated by digesting pRLT-2 with Ecl XI and Bam HI, gel eluting the approximately 10 kilobase pair (kb) linearized band, treating the fragment with Klenow and religating the fragment to generate a plasmid which encoded only NS3 (7606-9476 nt). pRLT-3 was subsequently digested with Xho I to generate a 1 kb fragment which was gel eluted, purified and ligated into Xho I digested and dephosphorylated pGEX-4T-2 (Pharmacia Biotech, Inc., Piscataway, New Jersey) to generate plasmid pT-1.

B. Construction of plasmid pT-3

The NS4A gene was amplified by PCR using reagents from an AmpliTaq kit (Perkin Elmer, Foster City, CA), with the following primers:

(1) SEQ ID NO:17 5'-GTGGCCCACCTGCATGCTAGCACCTGGGTGCTCGTT-3' and

17

(2) SEQ ID NO:18 5'-ATGAATTCAGCACTCTTCCATCTCATCGAA-3' and pBRTM/HCV
digested with Ecl XI as template DNA. The PCR product was cloned into vector pCR II
(Invitrogen, San Diego, CA) according to the manufacturer's instructions. The resulting vector
was digested with Sph I and Not I and the 209 base pair fragment containing the NS4A region

5   was gel purified. Plasmid pT-1 was then digested with Sph I and Not I and the 5.584 kb
fragment containing the NS3 portion was gel-purified and ligated to the 209 base pair SphI/Not
I fragment of NS4A to generate plasmid pT-3. Plasmid pT-3 was transformed into *E. coli*
strain JM109 (Promega, Madison, WI) for expression studies. Expression of plasmid pT-3
was shown (by SDS-PAGE gel electrophoresis) to produce a fusion protein of NH_2-GST-

10   NS3-NS4A-COOH of approximately 54.8 kD.

Example 2: Isolation and Purification of NS3/4A Fusion Protein

A.   Large-scale preparation of strain pT-3/JM109

15   Strain pT-3/JM109 was grown for large scale production as follows: A starter culture
(20 ml) containing Tryptone-Phosphate broth (TP/liter = 20 g tryptone, 15 g yeast extract, 8 g
NaCl, 2 g Na_2HPO_4, and 1 g KH_2PO_4), ampicillin (Amp, 100 ug/ml) and 0.2% dextrose was
inoculated with a single colony of pT-3/JM109 and shaken at 250 rpm at 37°C for 16 hours.
This culture was used to inoculate (at a 1:50 dilution) 1 liter of TP broth (including the named

20   antibiotics and dextrose) in a 2.8 Liter Fernback flask. The culture was shaken at 37°C for 2
hours, after which 1 mL of 100 mM IPTG was added and the culture shaken for an additional 2
hours. The culture was then aliquoted (250 ml aliquots) into Corning centrifuge bottles (Cat.
No. 25350-250) and the cells harvested by centrifuging at 2,000 rpm in a Sorvall GSA rotor at
4°C. The supernatant was discarded and the wet pellet weighed. The pellets were frozen in a

25   dry ice/ethanol bath and stored at -80°C until further use.

B.   GST purification of the GST/HCV fusion protein

        A 250 ml culture pellet was thawed and resuspended in 10 ml 1X STE (10 mM Tris,
pH 8.0, 150 mM NaCl, 1 mM EDTA). The cells were lysed with a French press cell two times

30   at 20 kpsi and the lysate placed into plastic 15 ml oakridge tubes. Triton X-100 (10% solution
in 1x PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.4) was
added to the lysate to a final concentration of 1.0% and the solution was mixed gently at room
temperature for 30 minutes. DTT (either solid or a 1 M stock) was added to a final
concentration of 20 mM and the solution again mixed gently at room temperature for 5 minutes.

35   The solution was placed on ice for 1 hour and then centrifuged at 14,000 rpm in a Sorvall
SA600 rotor for 20 min. at 4°C. The supernatant was carefully decanted into a second 15 ml
oakridge tube and then applied to a GST Sepharose column (2 mL, available from Pharmacia

18

Biotech, Inc., Piscataway, New Jersey) which had been equilibrated with 10 mL of 1x PBS containing 1% Triton X-100. The column was washed at room temperature with at least 100 to 150 mL of 1x PBS containing 1% Triton X-100. Elution buffer (10mL, having 50 mM Tricine, pH 8.0, 10 mM reduced glutathione, 5 mM DTT, and 1% Triton X-100) was then applied to the column, collected in fractions and assayed for protease activity in the manner described in Example 3 below. Fractions having enzyme activity were stored in elution buffer at -80°C until future use. Total protein content was later determined by detergent compatible Bradford assay (BioRad, Hercules, CA).

Example 3: Demonstration of HCV Protease Activity

A. Methodologies and Reagents

1. Assay Reagents: Fluorogenic peptide substrate having the sequence Ac-G-E(EDANS)-(ethylene glycol linker)-E-D-V-V-A-C-S-M-S-Y-(ethylene glycol linker)-K(Dabycl)-G-NH$_2$ (hereinafter termed "FPS-1") was synthesized and purified according to the procedure of E. D. Matayoshi et. al., Science 247: 954, 1990. FPS-1 is a cleavage substrate having a modified NS5A/5B cleavage junction, the modification being the substitution of amino acid A for the P2 amino acid C in SEQ ID NO:13. In FPS-1, the P1 amino acid is C and the P1' amino acid is S. Accordingly, proper cleavage of FPS-1 results in the following two products: Ac-G-E(EDANS)-(ethylene glycol linker)-E-D-V-V-A-C (SEQ ID NO:19) and S-M-S-Y-(ethylene glycol linker)-K(Dabycl)-G-NH$_2$ (SEQ ID NO:20).

2. Kinetics Assay: Kinetics assays (for determining fusion protein activity) were performed as 200 µL reactions containing 50 mM Tricine, pH 8.0; 30 % glycerol, 0.2% Triton X-100, 6 µM synthetic peptide substrate (pre-incubated as a 50 µM solution in 2 mM DTT for at least 30 minutes prior to use) with purified GST-NS3/4A fusion protein. As negative controls, GST protein, a fusion protein of GST-CMV protease or no protein were used in place of GST-NS3/4A fusion protein under otherwise identical reaction conditions. Assay mixtures were incubated at room temperature and the progress of the reaction monitored for up to 1 hour in a Titertek Fluoroskan II instrument (ICN Biomedicals, Huntsville, AL) with an excitation filter set at 335 nm and emission filter at 485 nm. Data was collected online with a Macintosh computer using DELTA SOFT II, version 4.0 (BioMetallics, Inc., Princeton, NJ). Nonlinear curve fitting was performed using KaleidaGraph (Synergy Software, Reading, PA).

In kinetic assays performed using the pMAL constructs (i.e. the "NS3 series") and described below in Example 4, the assays were performed essentially as described above with the modification that NS4A peptide also was added to the reaction.

3. Total Cleavage Assay: Total cleavage assays were performed in essentially the same manner as the kinetics assays with the following modifications: the reaction mixtures were

19

scaled up to 1 mL and incubated overnight at room temperature either with or without enzyme. Samples from the reactions were then injected onto an HPLC RP18 column and the products separated with an acetonitrile linear gradient of 15-35% in 40 minutes (i.e. 0.5% change per minute). The progress of the digested products was monitored by absorbance at 470 nm to detect the dabcyl moiety and fluorescence (excitation filter = 355 nm, emission filter = 470 nm) to detect the EDANS moiety.

B. Results:

1. Kinetics Analysis: As shown in FIG. 4, the amount of EDANS fluorescence increased over time when the FPS-1 substrate was incubated in the presence of purified GST-NS3/4A fusion protein under the conditions described above for the kinetics assay. In contrast, a cleavable substrate incubated with no protein, GST protein or a fusion protein of GST-CMV protease did not generate a measurable increase in fluorescence. (GST-CMV protease had been shown to be fully active in the kinetics assay using its authentic substrate (data not shown)).

2. Product Analysis: As shown in FIG. 5, when the total cleavage assay was performed in the absence of an NS3/4A fusion protein, a single absorbance peak (which comigrated with the major fluorescence peak) was seen at 39.8 minutes retension time (RT). Mass spectral analysis showed this peak to be intact substrate. When the assay was performed in the presence of an NS3/4A fusion protein, two peaks appeared with retention times of 8.7 minutes and 37.3 minutes (fluorescence and absorbance respectively) whereas the substrate peak at RT = 39.8 minutes diminished (see FIG. 5, 2-25 Hr time points). These results were consistent with the prediction that upon cleavage of the substrate, the internal quenching effect by the dabcyl moiety would be eliminated so that the N-terminal fragment would display substantial fluorescence intensity while the C-terminal fragment would display dabcyl absorption only. Peptide sequencing of the dabcyl containing C-terminal fragment showed it to have the sequence of S-M-S-Y; mass spectral analysis also confirmed the identity of each peak as cleavage products having expected molecular weights. These results demonstrate that a non-autocleavable NS3/4A fusion protein is biologically active and capable of cleaving a peptide substrate at the C-S scissile bond of a modified NS5A/5B cleavage junction.

Example 4: Construction of MBP-cut site-NS3/4A clones

Fusion proteins were generated and experiments performed to demonstrate that NS3/4A fusion proteins of the present invention have full *cis*-cleavage activity. Such proteins are envisioned for use to screen compounds which inhibit the protease activity and/or to study the protease substrate requirement by mutagenesis methods.

20

a. Construction and Purification of MBP-cut site-NS3/4A clones: HCV NS3 protein
[corresponding to amino acid positions 1-181 of SEQ ID NO:4 and designated as "NS3 series"
in FIG. 6(b)] as well as NS3/NS4A fusion protein (corresponding to SEQ ID NO:4 and
designated as."T3 series" in FIG. 6(a)] were cloned into pMAL-c2 vectors obtained from New
England BioLabs (Bevery, MA). HCV NS3 serine protease cleavage sites were also inserted
between the maltose binding protein (MBP) and the serine protease domain. The peptide
sequences of the cleavage sites were as follows: NS3/4A site: DLEVVT-STWV (amino acids 1-
10 of SEQ ID NO:10); NS4A/4B site: DEMEEC-SQHL (amino acids 1-10 of SEQ ID NO:11);
NS4B/5A site: ECTTPC-SGSW (amino acids 1-10 of SEQ ID NO:12); and NS5A/5B site:
EDVVCC-SMSY (SEQ ID NO:15). In each sequence shown above, the scissile bond is
indicated by a dash (-). An active site mutation also was generated in the 5A/5B cut site
construct (called pMAL-5AB-D81N) within the NS3 protein at amino acid position 81. At that
position, the Asp was mutated to Asn (and is referred to in FIG. 6(b) as D81N). To complete
the constructs, a six histidine tag was linked to the carboxy terminus of each of these fusion
proteins.

The constructs were transformed into *E. coli* JM109 bacteria. Synthesis of the fusion
proteins was induced using IPTG under standard conditions. Gene products were analyzed by
SDS-PAGE, Western blot analysis and MBP affinity purification. MBP fusion proteins were
purified using amylose resin (New England BioLabs) whereas the his-tagged polypeptides
were purified by Talon metal affinity resin (Clontech, Palo Alto, CA). Western analysis was
performed with anti-His antibody (Invitrogen, Carlsbad, CA) and visualized with an ECL
Western blotting analysis system (Amersham, Arlington Hights, IL). All procedures described
in this example were performed using standard molecular biology and biochemistry techniques
or according to manufacturer's instructions.

b. Results: When the whole cell lysate of JM109 containing the construct pMAL-23-
T3 was analyzed by SDA-PAGE, an overexpressed MBP fusion protein of 63.5 KD in size
was easily visualized by Coomassie blue staining. Protein purification carried out on this lysate
by amylose affinity chromatography also retrieved this 63.5 KD polypeptide. The fusion
protein demonstrated full NS3 serine protease activity in the peptide cleavage assay described
above (in Example 3).

Analysis of other proteins in the T3 series indicated that purified MBP-fusions were all
active in the peptide cleavage assay. However, the proteins showed different levels of self-
cleaving activity, the most susceptible being the fusion protein containing the 5A/5B site
followed by fusion proteins containing 4A/4B and 4B/5A sites. No indication of autocleavage
was observed in the pMAL-34-T3 construct.

The same analysis was also performed on the NS3 series of pMAL constructs.
Autocleavage activities were observed in the constructs containing the 4A/4B, 4B/5A and

5A/5B sites. No indication of self-cleavage was apparent in the pMAL-34-NS3 protein. The active site mutation D81N of pMAL-5AB-D81N abolished the protease activity. The self-cleavage activity followed same order as observed in the T3 series, i.e. 5A/5B > 4A/4B > 4B/5A. Addition of synthetic NS4A peptide in the incubation buffer containing full length

5   MBP fusion proteins stimulated the self-cleavage activities of the proteins with 4A/4B and 4B/5A sites to more than 60%. No stimulation was found with the MBP-5AB-NS3 fusion protein. These results are in agreement with other observations that 5A/5B cleavage is independent of NS4A whereas 4A/4B and 4B/5A sites require the addition of NS4A to effect efficient cleavage.

10

22

We claim:

1.   An isolated or purified polynucleotide comprising a nucleotide sequence (A) having a nucleotide sequence (B) or fragments thereof which encode hepatitis C virus NS3 protease and a nucleotide sequence (C) or fragments thereof which encode NS4A cofactor protein, wherein said nucleotide sequence (A) produces, upon expression, a non-autocleavable fusion protein of hepatitis C virus NS3 protease and hepatitis C virus NS4A cofactor protein which is biologically active.

2.   The polynucleotide of Claim 1 wherein said nucleotide sequence (B) is located upstream from said nucleotide sequence (C).

3.   The polynucleotide of Claim 1 wherein said biologically active fusion protein is capable of cleaving at least SEQ ID NO:16.

4.   The polynucleotide of Claim 1 wherein said nucleotide sequence (A) is SEQ ID NO:3.

5.   The polynucleotide of Claim 1 wherein said nucleotide sequence (B) encodes a biologically active domain of NS3 protease.

6.   The polynucleotide of Claim 5 wherein said nucleotide sequence (B) comprises from about nucleotide position 1 to about nucleotide position 543 of SEQ ID NO:1.

7.   The polynucleotide of Claim 1 wherein said nucleotide sequence (C) encodes a biologically active domain of NS4A cofactor protein.

8.   The polynucleotide of Claim 5 wherein said nucleotide sequence (C) comprises from about nucleotide position 1957 to about nucleotide position 1995 of SEQ ID NO:1.

9.   An expression vector comprising the polynucleotide of Claim 1 or Claim 4.

10.   The expression vector of Claim 9 further comprising an enhancer-promoter operatively linked to said polynucleotide.

11.   The expression vector of Claim 9 which is a pGEX vector.

23

12. A host cell transformed with the expression vector of Claim 9.

13. The host cell of Claim 12 that is a eukaryotic or prokaryotic cell.

14. A biologically active fusion polypeptide comprising hepatitis C virus NS3 protease and hepatitis C virus NS4A cofactor protein wherein said fusion protein is non-autocleavable.

15. The fusion protein of Claim 14 which is capable of cleaving at least SEQ ID NO:16.

16. The fusion protein of Claim 14 having SEQ ID NO:4.

17. The fusion protein of Claim 14 which is capable of cleaving a substrate comprising SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9.

18. A method for identifying an inhibitor compound of hepatitis C virus NS3 protease comprising the steps of:

(a) providing a reaction mixture having (i) a substrate wherein said substrate is capable of being cleaved by a hepatitis C virus NS3 protease acting alone or in combination with a hepatitis C virus NS4A cofactor protein, (ii) a non-autocleavable fusion protein of hepatitis C virus NS3 protease and hepatitis C virus NS4A cofactor protein which is biologically active and (iii) a compound of interest;

(b) incubating said reaction mixture; and

(c) determining the extent of cleavage of said substrate in said reaction mixture.

19. The method of Claim 22 wherein said said fusion protein has SEQ ID NO:4.

FIG. 1(a)

```
1    GCG CCC ATC ACG GCG TAC GCC CAG CAG ACG
     A   P   I   T   A   Y   A   Q   Q   T
31   AGA GGC CTC CTA GGG TGT ATA ATC ACC AGC
     R   G   L   L   G   C   I   I   T   S
61   CTG ACT GGC CGG GAC
     L   T   G   R   D

91   AAA AAC CAA GTG GAG GGT GTC CAG ATC GTG
     K   N   Q   V   E   G   V   Q   I   V
     TCA ACT GCT ACC CAA ACC TTC GCA ACG TGC
     S   T   A   T   Q   T   F   A   T   C
121  ATC AAT GGG
     I   N   G

151  GTA TGC TGG ACT GTC TAC TAC CAC GGG GCC
     V   C   W   T   V   Y   Y   H   G   A
181  GGA ACG AGG ACC TCA CCC AAG GGT CCT GTC
     G   T   R   T   S   P   K   G   P   V
211  ATC CAG ATG TAT
     I   Q   M   Y

241  ACC AAT GTG GAC GAC CTT GTG GGG TGG CCC
     T   N   V   D   D   L   V   G   W   P
271  GCT CCT CAA GGT TCC CGC TCC TTG ACA CCC
     A   P   Q   G   S   R   S   L   T   P
     TGC ACC TGC G
     C   T   C

301  TCC TCG GAC CTT TAC CTG AGG CAC GCC GAC
     S   S   D   L   Y   L   R   H   A   D
331  GTC ATT CCC GTG CGC CGG CGA GGT GAT AGC
     V   I   P   V   R   R   R   G   D   S
361  AGG GGT
     R   G

391  CTG CTT TCG CCC CGG CCC ATT TCC TAC CTA
     L   L   S   P   R   P   I   S   Y   L
421  AAA GGC GGT GGG CCG CTG CCC CCG GCA CAC
     K   G   G   G   P   L   P   P   A   H

451  GTG GGC CTA TTC AGG GCC GTG TGC ACC CGT
     V   G   L   F   R   A   V   C   T   R
481  GGA GTG GAC TTT ATC CCT GTG GAG AAC
     G   V   D   F   I   P   V   E   N
511

541  GAG ACA ATG AGA TCC CCG GTG GAC AAC TCC
     E   T   M   R   S   P   V   D   N   S
571  CCA CCA GCA GTG CCC CAG AGC TTC CAG GTG
     P   P   A   V   P   Q   S   F   Q   V

601  CAC CTG CAT GCT CCC GGC AGT GGT AAG AGC
     H   L   H   A   P   G   S   G   K   S
631  ACC AAG GTC CCG GCT GCA TAC GCA CAG GGC
     T   K   V   P   A   A   Y   A   Q   G
661  TAC AAG
     Y   K

691  TTG GTG CTC AAC CCC TCT GTT GCT ACG CTG
     L   V   L   N   P   S   V   A   T   L
721  GGC TTT GGT GCC TAC ATG TCC AAG TCC CAT
     G   F   G   A   Y   M   S   K   S   H
     GGG GTC GAT
     G   V   D

751  AAT ATC AGG ACC GGG GTG AGA ACA ATT ACC
     N   I   R   T   G   V   R   T   I   T
781  ACT GGC AGC AGC TAC ACC GGC AAG TTC CTT
     T   G   S   S   Y   T   G   K   F   L
811  GCC
     A
```

FIG. 1(b)

```
841
GAC GGC GGG TGC TCA GGA GGC GCT TAT GAC ATA ATA ATT TGT GAC GAG TGC CAC TCC ACG GAT GCC ACA TCC ATC
 D   G   G   C   S   G   G   A   Y   D   I   I   I   C   D   E   C   H   S   T   D   A   T   S   I
                                              871

901
TTG GGC ATC GGC ACT GTC CTT GAC CAA GCA GAG ACT GCG GGG GCG AGA TTG GTT GTT CTC ACT GCT ACC CCT
 L   G   I   G   T   V   L   D   Q   A   E   T   A   G   A   R   L   V   V   L   T   A   T   P
                            931

961
CCG GGC TCC GTC ACT GTG TCC CAT CCT AAC ATC GAG GAG GTT GCT CTG AGC ACC AAG AAG TGC GAC
 P   G   S   V   T   V   S   H   P   N   I   E   E   V   A   L   S   T   K   K   C   D
                        991

1021
GGC AAG GCT ATC CCC CTC GAG GTG ATC AAG AAG GGG AGA CAT CTC ATC TTC TGT CAC TCA GTC
 G   K   A   I   P   L   E   V   I   K   K   G   R   H   L   I   F   C   H   S   V
           1051

1081
GAG CTC GCC GCG AAG CTG GTC GCA TTG GGC ATC AAT GCC GTG GCC TAC TAC CGC GGA CTT GAC GTG
 E   L   A   K   L   V   A   L   G   I   N   A   V   A   Y   Y   R   G   L   D   V
           1141                     1171

1201
CCG ACC AAC GGC GAT GTT GTC GTC TCG GAT GCT CTG ATG ACT GGC TTT ACC GGC GAC TTC GAC
 P   T   N   G   D   V   V   V   S   D   A   L   M   T   G   F   T   G   D   F   D
           1231                     1261

1291
ATA GAC TGC AAC ACG TGT GTC ACT CAG GAT TTC AGC CTT GAC CCT TTT ATT GAG ACA ACC ACG
 I   D   C   N   T   C   V   T   Q   D   F   S   L   D   P   T   I   E   T   T   T
           1321

1351
CTC CCC CAG GAT GCT GTC TCC AGG ACT CAG GGC CGG AGG GGG AAG CCA GGG ATC TAC AGA TTT
 L   P   Q   D   A   V   S   R   T   Q   R   G   K   P   G   I   Y   R   F
           1381                     1411

1441
GTG GCA CCG GGG GAG CGG CCC TCC ATG TTC GAC TCG TCC GTC CTC TGT GAG TGC TAT GAC GCG GGC
 V   A   P   G   E   R   P   S   G   M   F   D   S   S   V   L   C   E   C   Y   D   A   G
           1471                     1501

1531
TGG TAT GAG CTC ATG CCC GAG ACT ACA GTT CGA CTA CGA TAC ATG AAC ACC CCG GGG CTT CCC
 W   Y   E   L   M   P   A   E   T   V   R   L   R   Y   M   N   T   P   G   L   P
                    1561

1591
CAG GAC CAT CTT GAA TTT TGG GAG GGC GTC TTT ACG GGC CTC ACC CAT ATA GAT GCC CAC TTT CTA TCC CAG ACA
 Q   D   H   L   E   F   W   E   G   V   F   T   G   L   T   H   I   D   A   H   F   L   S   Q   T
                                              1621
```

FIG. 1(c)

```
1651
AAG CAG AGT GGG GAG AAC TTT CCT TAC     1681  TAC CTG GTA GCG TAC CAA ACC GTG TGC GCT     1711  GCT CAA GCC CCT CCC
 K   Q   S   G   E   N   F   P   Y             L   V   A   Y   Q   A   T   V   C   A             A   Q   A   P   P

CCA TCG TGG GAC CAG ATG TGG AAG TGT TTG   1741  ATC CGC CTT AAA CCC ACC CTC CAT GGG CCA   1771  ACA CCC CTG CTA TAC
 P   S   W   D   Q   M   W   K   C   L           I   R   L   K   P   T   L   H   G   P           T   P   L   L   Y

1801
AGA CTG GGC GCT GTT CAG AAT GAA GTC ACC   1831  CTG ACG CAC CCA ATC ACC AAA TAC ATC ATG   1861  ACA TGC ATG TCG GCC
 R   L   G   A   V   Q   N   E   V   T           L   T   H   P   I   T   K   Y   I   M           T   C   M   S   A

GAC CTG GAG GTC ACG AGC ACC TGG GTG CTC   1891  GTT GGC GGC GTC GCT CTG GCC GCG TAT TGC   1921  CTG TCA
 D   L   E   V   T   S   T   W   V   L           V   G   G   V   L   A   L   A   Y   C           L   S

1951
ACA GGC TGC GTG GTC ATA GTG GGC AGG ATT         GTC TTG TCC GGG AAG CCG GCA ATT ATA CCT   2011  GAC AGG GAG GTT CTC
 T   G   C   V   V   I   V   G   R   I     1981  V   L   S   G   K   P   A   I   I   P           D   R   E   V   L

TAC CAG GAG TTC GAT GAG ATG GAA GAG TGC     2041
 Y   Q   E   F   D   E   M   E   E   C
```

FIG. 2

```
  1 GCGCCCATCA CGGCGGTACGC CCAGCAGACG AGAGGCCTCC TAGGGTGTAT AATCACCAGC CTGACTGGCC GGGACAAAAA  80
 81 CCAAGTGGAG GGTGAGGTCC AGATCGTGTC AACTGCTACC CAAACCTTCC TGGCAAGGTG CATCAATGGG GTATGCTGGA 160
161 CTGTCTACCA CGGGGCCGGA ACGAGGACCA TCGCATCACC CAAGGGTCCT GTCATCCAGA TGTATACCAA TGTGGACCAA 240
241 GACCTTGTGG GCTGGCCCGC TCCTCAAGGT TCCCGCTCAT TGACACCCTG CACCTGCGGC TCCTCGGACC TTTACCTGGT 320
321 TACGAGGCAC GCCGACGTCA TTCCCGTGCG CCGGCGAGGT GATAGCAGGG GTAGCCTGCT TTCGCCCCGG CCCATTTCCT 400
401 ACCTAAAAGG CTCCTCGGGG GGTCCGCTGT TGTGCCCCGC GGGACACGCC GGGTGTGCACC                       480
481 CGTGGAGTGA CCAAGGCGGT GGACTTTATC CCTGTGGAGA AACCATGAGA TCCCCGGTGT TCAGGGGCCGC            560
561 CTCCTCTCCA CCAGCAGTGC CCCAGAGCTT CCAGGTGGCC CACCTGCATG CTAGCACCTG GGTGCTCGTT GGCGGCGTCC 640
641 TGGCTGCTCT GGCCGCGGTAT TGCCTGTCAA CAGGCTGCGT GGTCATAGTG GGCAGGATTG TCTTGTCCGG GAAGCCGGCA 720
721 ATTATACCTG ACAGGGAGGT TCTCTACCAG GAGTTCGATG AGATGGAAGA GTGCTGA                           777
```

FIG. 3

```
  1 APITAYAQQT RGLLGCIITS LTGRDKNQVE GEVQIVSTAT QTFLATCING VCWTVYHGAG TRTIASPKGP VIQMYTNVDQ  80
 81 DLVGWPAPQG SRSLTPCTCG SSDLYLVTRH ADVIPVRRRG DSRGSLLSPR PISYLKGSSG GPLLCPAGHA VGLFRAAVCT 160
161 RGVTKAVDFI PVENLETTMR SPVFTDNSSP PAVPQSFQVA HLHASTWVLV GGVLAALAAY CLSTGCVVIV GRIVLSGKPA 240
241 IIPDREVLYQ EFDEMEECZ                                                                     259
```
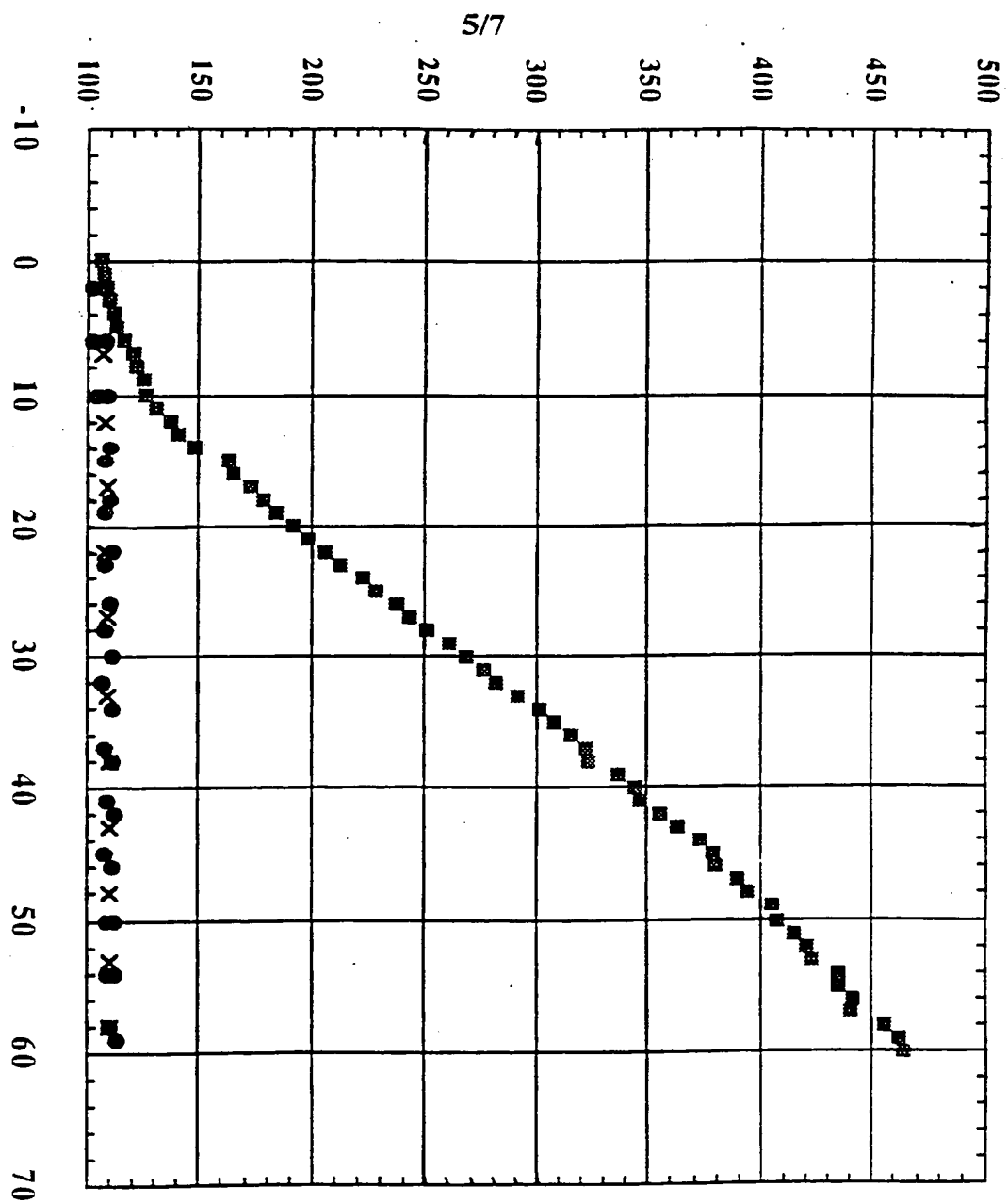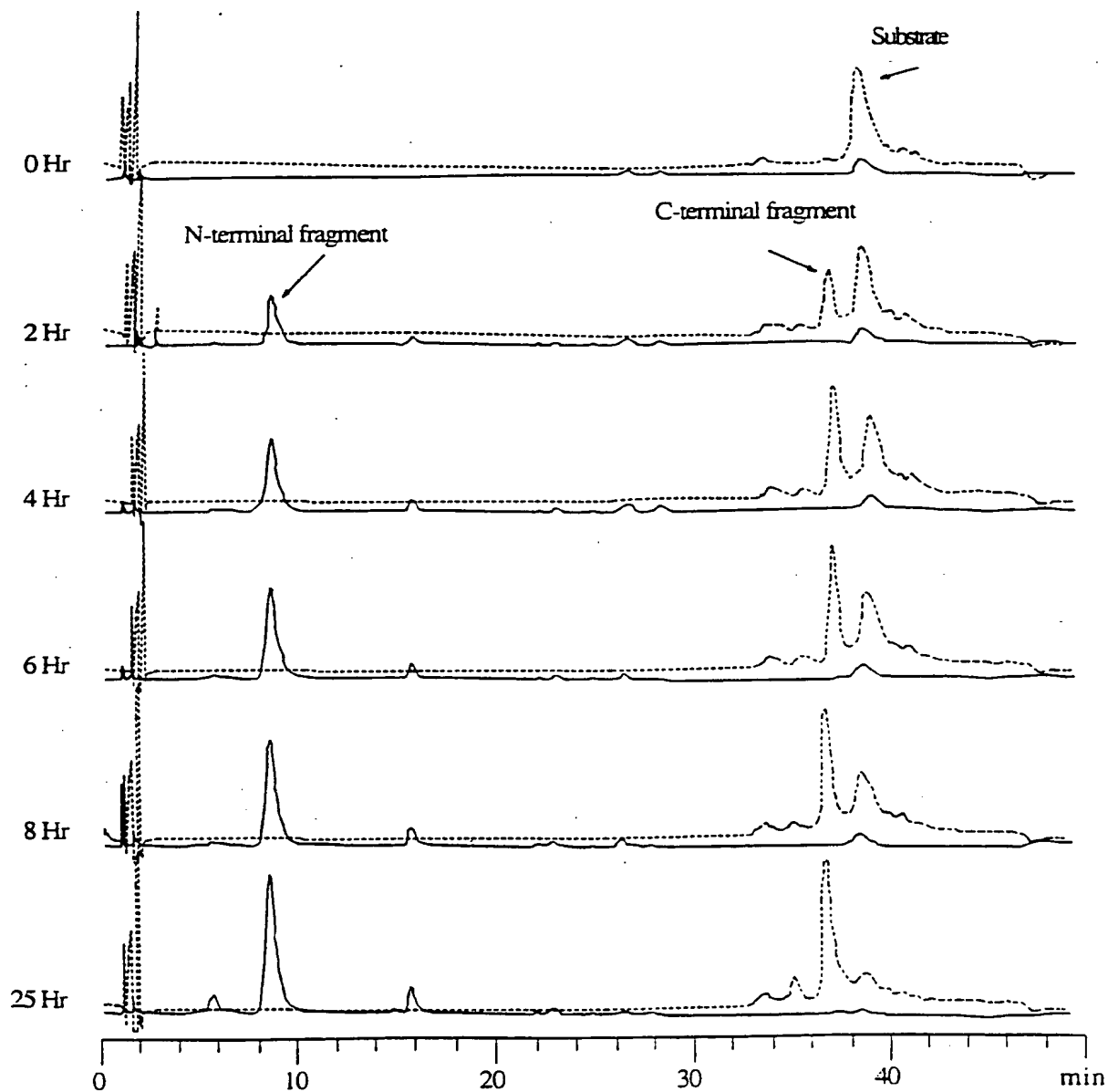
5/7

FIG. 4

6/7

FIG. 5

7/7

FIG. 6

a. T3 series

Constructs	Cut Site	Cut Site Sequence
pMAL-34-T3	3/4A	DLEVVT-STWV
pMAL-4AB-T3	4A/4B	DEMEEC-SQHL
pMAL-45-T3	4B/5A	ECTTPC-SGSW
pMAL-5AB-T3	5A/5B	EDVVCC-SMSY

b. NS3 series

pMAL-34-NS3	3/4A	DLEVVT-STWV
pMAL-4AB-NS3	4A/4B	DEMEEC-SQHL
pMAL-45-NS3	4B/5A	ECTTPC-SGSW
pMAL-5AB-NS3	5A/5B	EDVVCC-SMSY

pMAL-5AB-D81N	5A/5B	EDVVCC-SMSY

SUBSTITUTE SHEET (RULE 26)